Comparative analysis of breast cancer detection using K-means and FCM & EM segmentation techniques

V. Nagi Reddy[1,*], P. Subba Rao[2]

1. VFSTR Deemed to be University, Vadlamudi, Guntur, India

2. Dept of IT, VFSTR Deemed to be University, Vadlamudi, Guntur, India

nagireddy547@gmail.com

ABSTRACT. *About this study, we would like to existent breast cancer detection revealing procedures, established on the conservative a new spatial fuzzy-technique and K-means technique investigation of breast images. Although, the K-means was previously utilized in breast segmentation of image, along with segmentation of image at overall, this miss the mark to exploit the robust spatial association amongst neighbouring pixels. Spatial fuzzy C-means (sfcm's) procedure, that is exploit the evidence of spatial accurately and generate extraordinary breast image segmentation. To check the segmentation performance of spatial fuzzy C-means, K-means and expectation maximization methods, we have used 5 ground truth images. The outcomes of segmentation that are demonstrated extra precise segmentation with the sfcm's matched with that of K-means and expectation and maximization methods are offered statistically and graphically.*

RÉSUMÉ. *À propos de ces études, nous souhaitons mettre en place des procédures révélatrices de la détection du cancer du sein, qui ont été établies sur la base d'une nouvelle technique d'investigation par la technique floue spatiale et par la méthode K-moyennes des images MRI du sein. Bien que le K-moyennes ait déjà été utilisé dans la segmentation de l'image par MRI du sein, ainsi que dans l'ensemble, il manque la cible pour exploiter l'association spatiale robuste entre les pixels voisins. La procédure de C-moyennes flous spatials (SFCM), c'est-à-dire exploiter de manière précise la preuve spatiale et générer une extraordinaire segmentation des images du sein. Pour vérifier les performances de segmentation de C-moyennes flous spatials, de K-moyennes et des méthodes d'Espérance et de Maximisation, nous avons utilisé 5 images de vérité au sol. Les résultats de la segmentation, démontrés par une segmentation extrêmement précise avec les méthodes SFCM, sont comparés à ceux des méthodes K-moyennes et des méthodes d'Espérance et de Maximisation sont proposés de manière statistique et graphique.*

KEYWORDS: *SFCM, mammogram image, fuzzy, K-means, Em algorithm.*

MOTS-CLÉS: *SFCM, image MRI, Flou, K-moyennes, algorithme EM.*

DOI:10.3166/ISI.23.6.173-187 © 2018 Lavoisier

1. Introduction

In image, to pinpoint precise entity and limitations, normally segmentation of image is utilized. The consequence of image segmentation is a set of areas that jointly protect whole image, or a set of outlines take out from the image. To excerpt precise and miniscule evidence from composite images is significant procedure of Segmentation. This has comprehensive solicitation within the arena of science of medical. It may divide an image into commonly limited and fatigued areas such that respectively area of concern is spatially adjoining and pixels inside the area are uniform rendering to a predefined benchmark, through the presentation of image segmentation. The nature of breast images are difficult and inhomogeneous, renovating that inhomogeneous into homogeneous is actual challenging. This is actual challenging to envisage the configuration of the image for radiologists or urologists, without image segmentation. Here after, nearby is requisite of spontaneous segmentation methods on mammogram images of breast. The idea of spontaneous segmentation of breast is the compactness of the study. Using CT scan or mammogram scan, anatomy of the Breast can be observed. For the whole procedure, the mammogram scanned image is occupied in this study. For identification, the mammogram scan is extra relaxed than CT scan.

Scattering of existing tumor to another part of the body and grown as its own is called secondary tumor. Certainly, tumors are mass and malignant. Locating the malignant tumor is a little bit difficult whereas mass tumor detection is easy. For breast tumor detection, there are different types of algorithms used.We have analysed diverse natures of algorithms for the recognition of breast tumor using this study. Using Mat Lab, these algorithms are established. For the improvement and accomplishment, tt is stress-free. The approaches for tumor discovery and particular position of the tumor with precise borders is originate out lastly.

1.1. Literature review

To obtain high-resolution images of the breast, the mammogram is an extremely standard compared to Computed Tomography (CT). Normally, breasts are unprotected of advanced concentration of Fluoride (F) of the intake water, in many methods, then any additional human fleshy tissue. To provision the identification of different breast diseases such as breast stone, breast cancer, breast failure, polycystic breast disease, hematuria, nephritis, glomerulonephritis. We requisite the investigation of the spatial dissemination of those soft tissue.

Formerly, nearly of the approaches like-manual interactive threshold, slice editing and region painting are utilized for segmentation of medical image, that can be determined by graphical communication on human to describe areas of attention. The approaches of dissimilar for segmentation of image are confidential into 4 leading groups by yang *et al*. (2007) Threshold section increasing and edge based methods are measured as traditional procedures. Origination below statistical technique is the Maximum-Likelihood Classifier (MLC). Fundamentally, these procedures are controlled and be determined by the previous model with its constraints. Since the previous few years, approximately original approaches of segmentation are presented, which might be categorized as Statistical procedures.

For segmenting 3D medical images using a probabilistic administered moderation method was functional by Deighton *et al*. (2003). It will deliver the usage of indications to prime the segmentation. These signals are noticeable by nearly of the limitations such as standard deviation and mean. (Evangelin, Jensly and Suresh, Padma, 2015) worked on a 2D model of segmentation of the full breast. Used normalised gradients and a Mahalanobis distance from the time courses of the segmented regions to a training set for supervised segmentation.(Mahdi Marsousi, Konstantinos N. Plataniotis, Stergios Stergiopoulos, 2017) applied shape-to-volume registration, based on a new similarity metric, to detect the breast shape by fitting the 3-D shape model on 3-D ultrasound volumes. Fitted shape model is used to initialize and evolve a new level-set function, called complex-valued rational level-set with shape prior, to segment the breast's shape. By the standard fuzzy clustering algorithm, the areas of irregularity of the images are not accurately segmented (Chuang *et al*., 2006; Indah *et al*., 2011).

2. Methods and materials

2.1. Algorithm K-means

It is an exploratory technique and conservative, that provides improved productivity in clustering of data. This clustering is a cooperating method". This method assembly the evidence by frequently outcome of the statistical mean value for separately cluster in the group with adjacent mean after segmenting the image through categorizing each pixel. It stages apprehensive in the K-means procedure which are specified below:

a. Handpicked K primary clusters z1 (l), z2 (l),..........zK (l)

b. Kth recursive stage, proceeds the samples amongst k clusters specified beneath x€Cj(k) if $||x-zj(k)||<||x-zi(k)||$ For i = 1,2,..........,k, i≠j, wherever, Cj(k) indicates the payment of samples those cluster center is zj(k)

c. Achieve the fresh cluster centers zj (K+1), j = 1, 2,............, k, so that the Euclidean distance from points in Cj (K).Thus, ground-breaking cluster is specified by:

$Zi(k+1)= \frac{1}{N_j}\sum x\varepsilon Cj \;\; Xj = 1, 2, \ldots\ldots k$. where Nj is the quantity of examples in Cj(k).

If zj(k+1), j = 1 ,2 ,...k, the calculation will be congregated at the finale else go to stage 2.

Afterward breast mammogram image segmentation, it may catch out borders of the image with canny edge detection. Afterward, this catch out marking of the image and lastly precise position of the tumor within the image.

Benefits with K-means procedure:

- It will have all times K clusters
- In each cluster it will have all times at least one item

- Clusters do not overlay and are non- hierarchical.

Restrictions with K-Means procedure:

- This will be at all times ended
- At all times coming back K digit of clusters
- This is thoughtful to noise
- This necessitates extraordinary arrangement for minimization of energy.

2.2. Fuzzy C-means clustering (FCM)

Fuzzy-C Means stays spontaneous method which is efficiently utilized in MRI images due to objective documents and image segmentation. Fuzzy C-Means method tags the pixels with separate groups of values of data interested in bunches. Procedures which uses fuzzy segmentation retain extra information from the inventive images than compact techniques of segmentation. Fuzzy C-Means is established on classification of fuzzy pixel into precise areas, those fits into one course of pixels. FCM permits pixels to suitable onto numerous lessons taking unpredictable membership degrees. This kind of investigation provides humble presentation with a quantity of segmentation with real time data that is utilized frequently cutting-edge MRI.

Fuzzy C-means (FCM), a method of barrier into single slice with evidence from double or extra portions. Objective function minimization is specified through:

$$J=\sum_{j-1}^{N}\sum^{c}\ i=1\mu_{ij}^{m}\ \| X_j - V_i \|^2$$

Now, uij designates the pixel of membership function xj inside the ith cluster, vi specifies ith cluster centre and m is the ambiguity rate that is reserved as 2.

When pixel is nearby to centroid, the cost function is reduced who has maximum values of function of membership and which has least standards of bias function are assign to distant pixel commencing the centroid. Therefore, the function of membership provides pixel probability goes to specific cluster. The procedure depends on the extent of space among pixel and separable group taking the centroid through the domain of feature of the picture. The grade of function of membership and cluster centres stay produced with new values for each time by utilizing calculation:

$$\mu_{ij}=\frac{1}{\sum_{k-1}^{c}\left(\frac{\|x_j-v_i\|}{\|x_j-v_k\|}\right)^{\frac{2}{m-1}}} \text{ And } V_j=\frac{\sum_{j=1}^{N}u_{ij}^{m}x_j}{\sum_{j=1}^{N}U_{ij}^{m}}$$

Now uij Є [0, 1]. During early phase each bunch center of FCM meets to a definite vi, that provide burden topic of the function of cost. Fact of conjunction stands recognized with distinguishing variations with function of membership through cluster centroid at 2 repetition phases. An important piece of MRI image with together pixels take similar type standards and the possibility that the similar cluster is vast however, these relationship of spatial is important on clustering, this is not functional on a traditional FCM procedure.

2.3. Spatial FCM

Preparation of pixels in spatial field stands specified toward grows spatial domain evidence for example: $h_{ij}=\sum_{K \in NB}(X_j)\, u_{ik}$

Anywhere, NB (xj) symbolizes rectangular kernel centred with pixel xj in the image. A 3*3 kernel is executed for this technique. Now, the function hij signifies coincidental of receiving the pixel xj fit into ith cluster. The hij of a pixel on behalf of a cluster stands enormous once common of its neighbourhood partaking the alike cluster. The hij is announced interested in relationship grade for example charts:

$$U_{ij} = {u_{ij}^{p} h_{ij}^{q}} \Big/ {\sum_{k=1}^{c} u_{kj}^{p}\, h_{kj}^{q}}$$

Two comparative parameters are used as p and q. Those constraints consolidate the comparative significance of together selection functions. Using the part of homogeneity, the h_{ij} reinforce the actual association and the clustering consequences continue persistent. Nevertheless, with the case of a pixel through noise, these calculation shrinks the worth of a cluster with noise using the tags of its cover pixels. Therefore, not matching confidential pixels with area of noise or incorrect blobs may be forbidden. It is a double based method. In the initial repetition, the function of participation in the domain of frequency is premeditated. In the succeeding repetition, degree of the function of membership of apiece pixel is plotted to the domain with pixel and the h_{ij} is at that time assessed. The FCM repetition remains through a renewed function of membership till resolution congregates. The repetition pauses once the transformation among dual cluster working extreme for double consecutive repeation procedure is a smaller amount of a threshold (Є=0,02). Lastly, the resolution is touched, defuzzification is executed to allocate pixel to a precise cluster where the degree function is utmost.

Afterward breast mammogram image segmentation, it may catch out borders of the image with canny edge detection. Afterward, this catch out marking of the image and lastly precise position of the tumor within the image.

FCM procedure benefits:

- This delivers extra meticulous statistics than typical “rigid” grouping
- In conventional process, the classes number is less than unsupervised classification

FCM procedure restriction:

- It is challenging to select the constraints p and q for different breast MRI image.
- Challenging the usages of outliers.

2.4. Expectation and maximization algorithm

The EM algorithm leads to get having difficulties less than K-means algorithm.

The steps involved in EM algorithm

EM algorithm is performing with in two steps they are.

The (E) step:

With the help of currently using estimated parameter vector Φ, we can discover the expected value of Z_{ik}: $Z_{ik}^{t} = \frac{P_k^t G(x_i|\theta_k^t)}{f(x_i|\varphi^t)}$

From the above equation the posterior probability given xi is Zik,xi obtained from class k.

The matrix NxK is Z=[Zik] of the posterior probability satisfies the constraints, (0≤Zik≤1, $\sum kZik = 1, \sum iZik > 0{,}1 \leq i \leq N, 1 \leq k \leq K$).xi is the value of the pixel i.

The (M) step:

In this step, data expectation is shown below:

$$\mu_k^{(t+1)} = \frac{\sum_{i=0}^{N} Z_{ik}^t x_i}{\sum_{i=1}^{N} Z_{ik}^t}$$

$$\sigma^2 k^{t+1} = \frac{\sum_{i=1}^{N} Z_{ik}^t (x_i - \mu^{t+1})^2}{\sum_{i=1}^{N} Z_{ik}^t} \; P_K^{t+1} = \frac{\sum_{i=1}^{N} Z_{ik}^t}{N}$$

The parameters Φ is fixed, and the class probabilities are Z=[Z_{ik}].

3. Investigational outcomes

With MATLAB modelling, breast mammogram images are deliberated to authenticate the projected procedure. It is produced with Fuzzy C-means algorithm. Algorithm performance stays matched using K-Means and EM methods in relations of time complexity, noise and Quality metrics of individually breast MR images.

The performance of above segmentation algorithm is evaluated by using objective evaluation criteria like Jaccard index and volumetrity using formulas

$$\boldsymbol{JC} = \frac{|\boldsymbol{X} \cap \boldsymbol{Y}|}{|\boldsymbol{X} \cup \boldsymbol{Y}|} = \frac{\boldsymbol{a}}{\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c}}$$

$$\mathrm{VC} = 1 - \frac{||\mathrm{X}| - |\mathrm{Y}||}{|\mathrm{X}| + |\mathrm{Y}|} = 1 - \frac{\mathrm{b} - \mathrm{c}}{\mathrm{za} + \mathrm{b} + \mathrm{c}}$$

$$GCE = (S, S') = \frac{1}{N} \min\{\sum LRE(S, S', x_i), \sum LRE(S', S, x_i)\}$$

Where, LRE=$\frac{|C(S,x_i)\backslash C(S',x_i)|}{|C(S,x_i)|}$ S and S' are segment classes and x_i is the pixel.

$VOI(X,Y) = H(X) = H(Y) - 2I(X;Y)$ Here X and Y are two clusters

$PRI(S_t,\{S\}) = \frac{1}{N}\sum_{i,j,i<j}[I(l_i^{st} = l_j^{st})P_j + I(l_i^{st} \neq l_j^{st})(1-P_j)]$

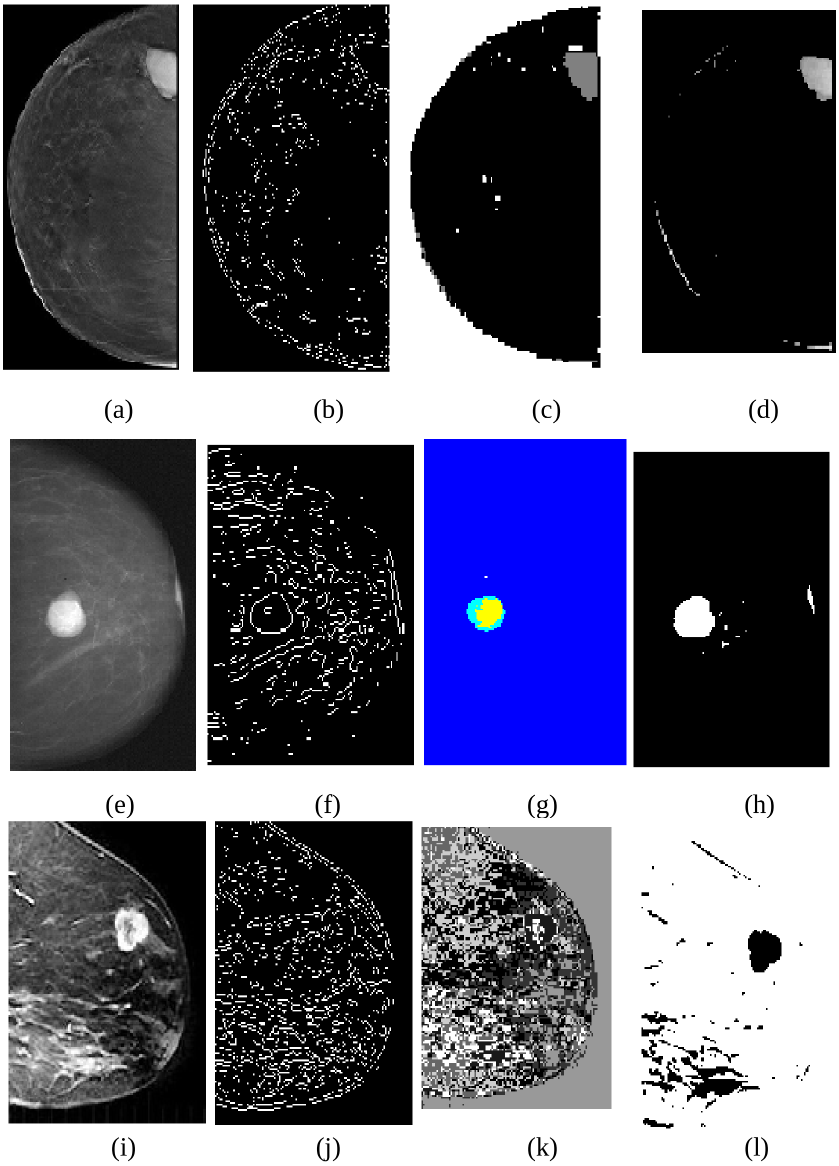


Figure 1. (a), (e), (i) are the original images the remaining are the images with region, label, tumor identification with respectively k-mean, EM, FCM algorithms

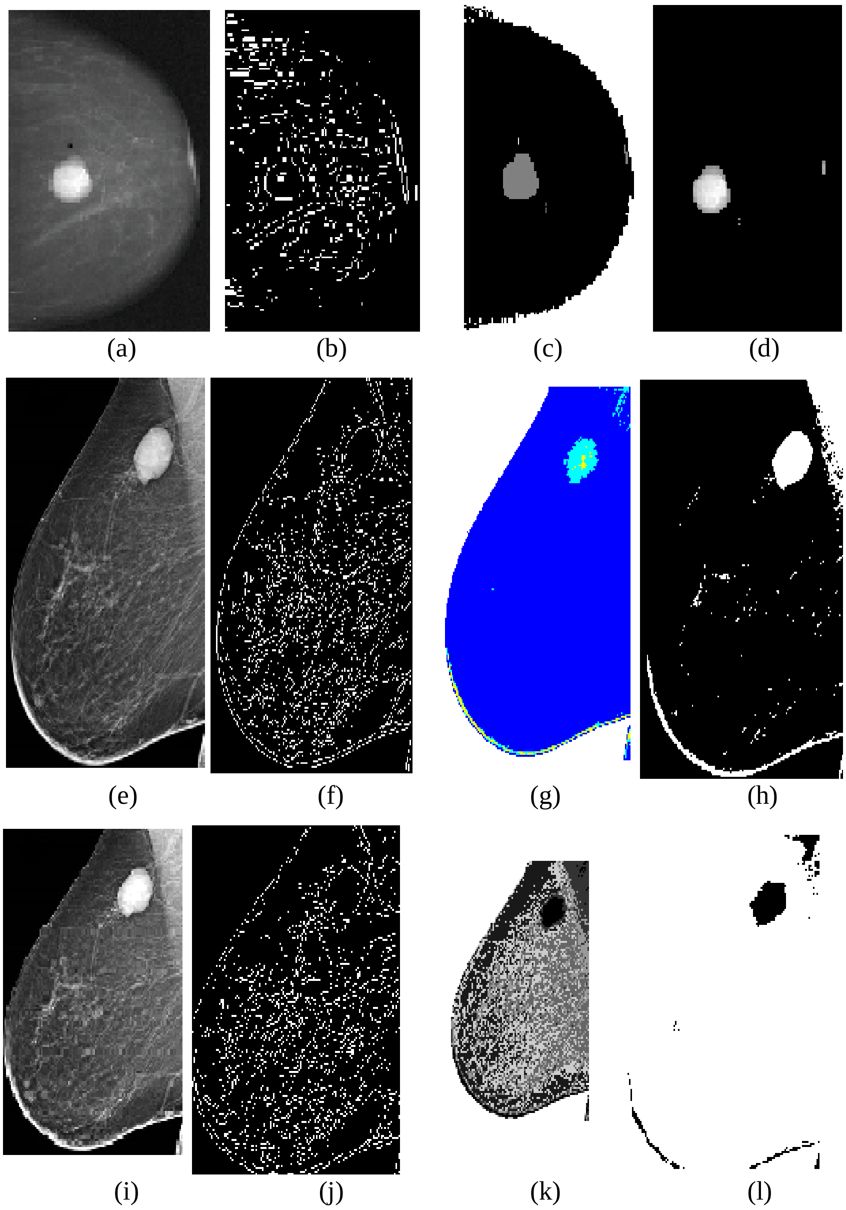
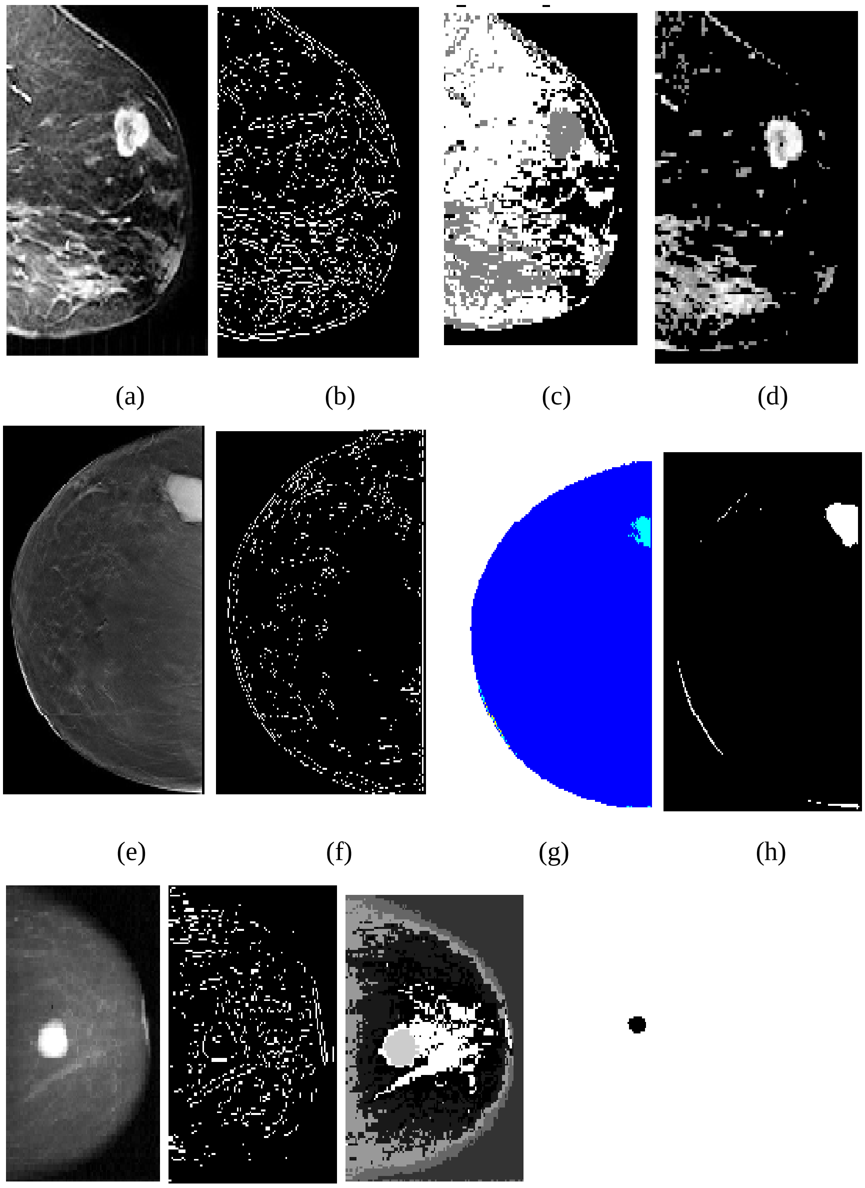


Figure 2. (a), (e), (i) are the original images the remaining are the images with region, label, tumor identification with respectively k-mean, EM, FCM algorithms



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 3. (a), (e), (i) are the original images the remaining are the images with region, label, tumor identification with respectively k-mean, EM, FCM algorithms

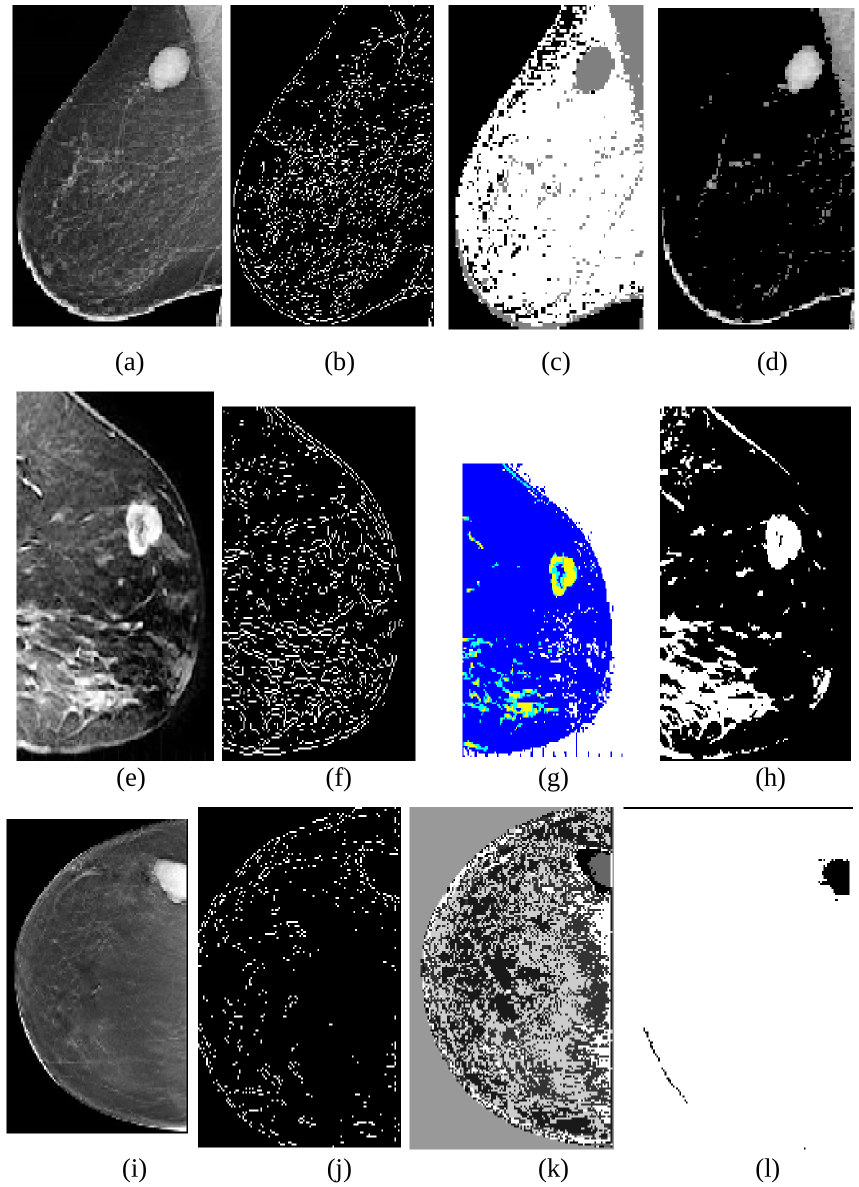


Figure 4. (a), (e), (i) are the original images the remaining are the images with region, label, tumor identification with respectively k-mean, EM, FCM algorithms

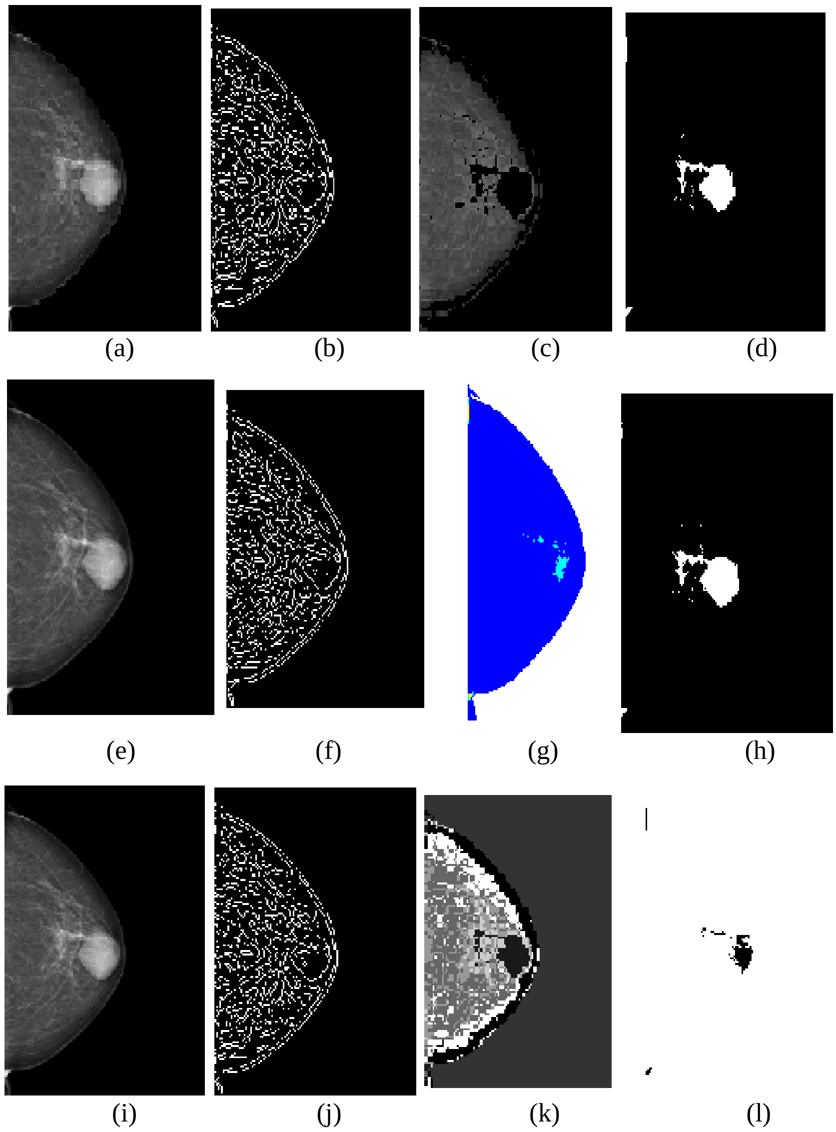


Figure 5. (a), (e), (i) are the original images the remaining are the images with region, label, tumor identification with respectively k-mean, EM, FCM algorithms

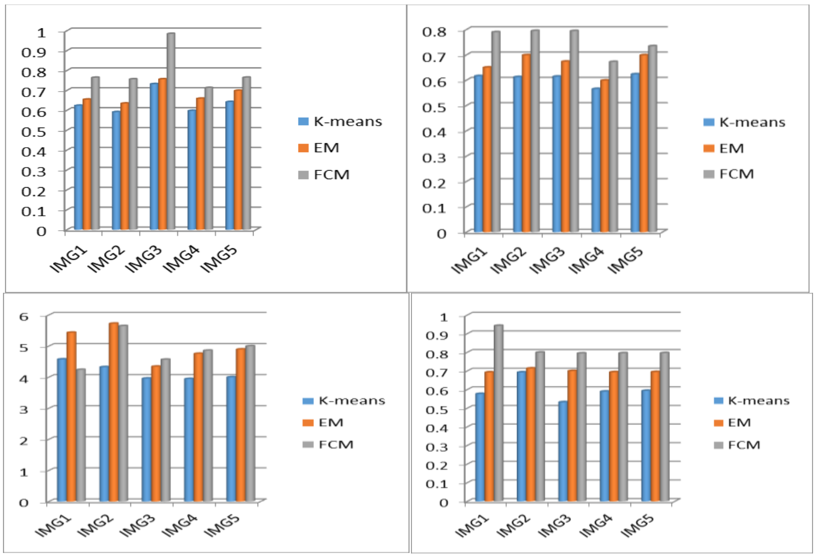


Figure 6. Jaccard Coefficient (JC), Volumetric Similarity (VS), Variation of Index (VOI), Global Consistency Error (GCE) for above five different breast images by using segmentation techniques

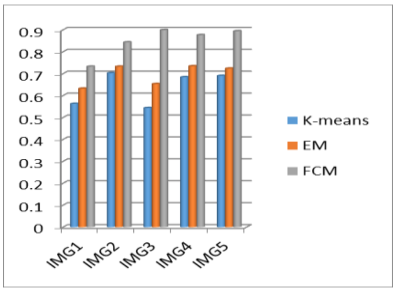


Figure 7. Probability random index for above five different breast images by using segmentation techniques

4. Discussions

From the above Figure1(b), (c) and (d) are relating K-Means procedure for these breast mammogram images outcomes therefore achieved are presented separately. This is witnessed that the performance of the procedure progressively worsens through growth by power of noise. The outcomes gained with using Fuzzy C-Means technique in this breast mammogram image is planned separately by Figure (i), (j), (k) and (l). The outcomes also gained with Expectation Maximization algorithem for breact mammogram images are arranged separately by Figure (e), (f), (g) and (h). Here all algorithms performance progressively worsens due to upturn with power of noise, nevertheless, segmentation outcome is extra precise plus satisfactory which equally matched as achieved with the K-Means & EM methods.

Table 1. This Below table shows the results of quality metrics by using K-mean, EM, FCM methods of five dissimilar breast mammogram images

Picture	Quality metrics	K-means	EM	FCM
Img1	JC	0.6235	0.6545	0.7643
	VS	0.6162	0.6502	0.7898
	VOI	4.5731	5.4321	4.2364
	GCE	0.5765	0.6923	0.9423
	PRI	0.5625	0.6321	0.7323
Img2	JC	0.5913	0.6345	0.7565
	VS	0.6123	0.6989	0.7948
	VOI	4.3245	5.7235	5.6453
	GCE	0.6923	0.7135	0.7989
	PRI	0.7023	0.7323	0.8432
Img3	JC	0.7321	0.7564	0.9847
	VS	0.6142	0.6734	0.7943
	VOI	3.9569	4.3412	4.5647
	GCE	0.5321	0.6998	0.7943
	PRI	0.5436	0.6531	0.8993
Img4	JC	0.5979	0.6589	0.7132
	VS	0.5654	0.5989	0.6723
	VOI	3.9435	4.7542	4.8543
	GCE	0.5894	0.6932	0.7954
	PRI	0.6843	0.7342	0.8764
Img5	JC	0.6423	0.6989	0.7649
	VS	0.6234	0.6987	0.7345

	VOI	3.9984	4.8937	4.9964
	GCE	0.5933	0.6943	0.7963
	PRI	0.6903	0.7234	0.8943

That K-Means technique contributes 68 toward 70% of correctness now altogether the circumstances with 5 breast mammogram images where as Fuzzy C-Means provides 80 toward 90% of exactness in altogether the circumstances with 5 breast mammogram images correspondingly.

The time complexity for individually to these approaches are too strongminded. It is observed that the time complexity drops for the event of Fuzzy C-Means procedure equally matched toward that of K-Means technique. The conforming values are correspondingly 0.6206, 0.92000 sec.

5. Conclusion

This presentation with diverse approaches for segmentation of image with breast mammogram is offered for these revisions. Projected Fuzzy C-Means procedure produce comparatively precise outcomes for all the circumstances of breast mammogram image segmentation. Additional, the projected procedure congregates quicker than K- means approaches.

References

Abo-Eleneen Z. A., Gamil A. (2013). A novel statistical approach for detection of suspicious regions in digital mammogram. *Journal of the Egyptian Mathematical Society*, Vol. 21, No. 2, pp. 162-168. https://doi.org/10.1016/j.joems.2013.02.002

Abuchaiba I. S. I., Elfarra B. K. (2013). New feature extraction method for mammogram CAD diagnosis. *International Journal of Signal Processing,* Vol. 6, No. 1.

Basheer N. M., Mohammed M. H. (2013). Segmentation of breast masses in digital mammograms using adaptive median filtering and texture analysis. *Int. J. Recent Technol. Eng. (IJRTE),* Vol. 2, No. 1, pp. 2277-3878.

Behrenbruch C., Marias K., Armitage P. (2001). Prone-supine breast MRI registration for surgical visualisation" medical imaging understanding and analysis. *Unv of Birmingham.* http://www.cs.bham.ac.uk/research/proceedings/miua2001/

DeWyngaert J. K., Noz M. E., Ellerin B., Kramer E. L., Maguire Jr G. Q., Zeleznik M. P. (2004). Procedure for unmasking localization information from prostascint scans for prostate radiation therapy treatment planning. *Int J Radiat Oncol Biol Phys*, Vol. 60, No. 2, pp. 654-662. https://doi.org/10.1016/j.ijrobp.2004.05.034

Dinsha D. (2014). Breast tumor Segmentation and classification using SVM and Bayesian from thremogram images. *Unique Journal of Engineering and Advanced Sciences,* Vol. 2, pp. 147-151.

Eubank W. B., Mankoff D. A., Vesselle H. J., Eary J. F., Schubert E. K., Dunnwald L. K., Lindsley S. K., Gralow J. R., Austin-Seymour M. M., Ellis G. K., Livingston R. B. (2002). Detection of locoregional and distant recurrences in breast cancer patients by using FDG PET. *Radio Graphics*, Vol. 22, No. 1, pp. 5-17. https://doi.org/10.1148/radiographics.22.1.g02ja055

Hathaway P. B., Mankoff D. A., Maravilla K. R., Austin-Seymour M. M., Ellis G. K., Gralow J. R., Cortese A. A., Hayes C. E., Moe R. E. (1999). Value of combined FDG PET and MR imaging in the evaluation of suspected recurrent local-regional breast cancer: Preliminary experience. *Radiology*, Vol. 210, pp. 807-814. https://doi.org/10.1148/radiology.210.3.r99mr43807

Inderpal S., Dinesh K. (2014). A review on different image segmentation techniques. *IJAR*, Vol. 4.

Khokher M. R., Ghafoor A., Siddiqui A. M. (2012). Image segmentation using multilevel graph cuts and graph development using fuzzy rule-based system. *IET Image Processing*, Vol. 7, No. 3, pp. 201-211. https://doi.org/10.1049/iet-ipr.2012.0082

Liberman L., Morris E. A., Lee M. J., Kaplan J. B., LaTrenta L. R., Menell J. H., Abramson A. F., Dashnaw S. M., Ballon D. J., Dershaw D. D. (2002). Breast lesions detected on MR imaging: Features and positive predictive value. *AJR*, Vol. 179, No. 1, pp. 171-178. https://doi.org/10.2214/ajr.179.1.1790171

Maguire Jr G. Q., Jaeger J., Farde L., Noz M. E. (1987). Use of graphical techniques for error evaluation. *Journal of Medical Systems*, Vol. 11, pp. 277-286. https://doi.org/10.1007/BF00994013

Maguire Jr G.Q., Noz M. E., Rusinek H., Jaeger J., Kramer E. L., Sange R J. J., Smith G. (1991). Graphics applied to image registration. *IEEE Computer Graphics and Applications*, Vol. 11, No. 2, pp. 20-28. https://doi.org/10.1109/38.75587

Noz M. E., Maguire Jr G. Q., Zeleznik M. P., Kramer E. L., Mahmoud F., Crafoord J. (2001). A versatile functional-anatomic image fusion method for volume data sets. *Journal of Medical Systems*, Vol. 25, pp. 297-307. https://doi.org/10.2147/jms.32412

Pfluger T., Vollmar C., Wismüller A., Dresel S., Berger F., Suntheim P., Leinsinger G., Hahn K. (2000). Quantitative comparison of automatic and interactive methods for MRI-SPECT image registration of the brain based on 3-dimensional calculation of error. *Journal of Nuclear Medicine*, Vol. 41, pp. 1823-1829.

Schettino C. J., Noz M. E., Kramer E., Taneja S., Lepor H. (2000). Impact of image fusion of the 111 in Capromab pendetide with MR or CT in patients with recurrent prostate CA. *Journal of Nuclear Medicine*, Vol. 42, pp. 294P. https://doi.org/10.2874/jnm.234714

Shanthi S., Bhaskaran V. M. (2014). Modified artificial bee colony based feature selection: A new method in the application of mammogram image classification. *International Journal of Science, Engineering and Technology Research*, Vol. 3, No. 6.

Shraddha T., Krishna K., Singh B. K., Singh R. P. (2012). Image segmentation: A review. *International Journal of Computer Science and Management Research*, Vol. 1, No. 4.

Teifke A., Hlawatsch A., Beier T., Werner Vomweg T., Schadmand S., Schmidt M., Lehr H., Thelen M. (2002). Undected malignancies of the breast: Dynamic contrast-enhanced MR imaging at 1.0 T. *Radiology*, Vol. 224, pp. 881-888. http://dx.doi.org/10.1148/radiol.2243010547